

The structure and evolution of language

Principles of Complex Systems
Course CSYS/MATH 300, Fall, 2009

Prof. Peter Dodds

Dept. of Mathematics & Statistics
Center for Complex Systems :: Vermont Advanced Computing Center
University of Vermont



Licensed under the *Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License*.

Language

Meanings
Irregular verbs
References

Frame 1/15



Outline

Meanings

Irregular verbs

References

Language

Meanings
Irregular verbs
References

Frame 2/15



Word meanings

Preliminary findings on word frequency and number of meanings

- ▶ Corpus: 10,000 most frequent words from Project Gutenberg
- ▶ # meanings for each word estimated using dictionary.com (田)
- ▶ Friends: perl, regular expressions, wget.

Language

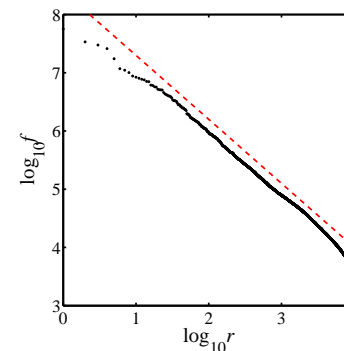
Meanings
Irregular verbs
References

Frame 3/15

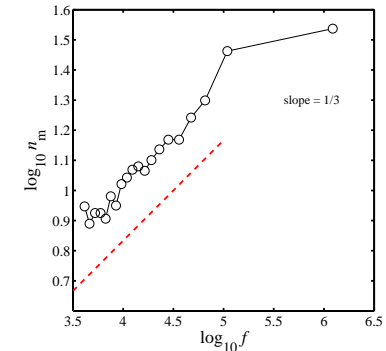


Word meanings

A.



B.



- A.** Word frequency versus rank, slope $\alpha \sim -1.1$ corresponds to a frequency distribution with $\gamma \sim 1.9$.
- B.** Relationship between average number of meanings and average frequency (bins are by rank, with each circle representing 500 words). Slope of $1/3$ lower than Zipf's $1/2$ [3].

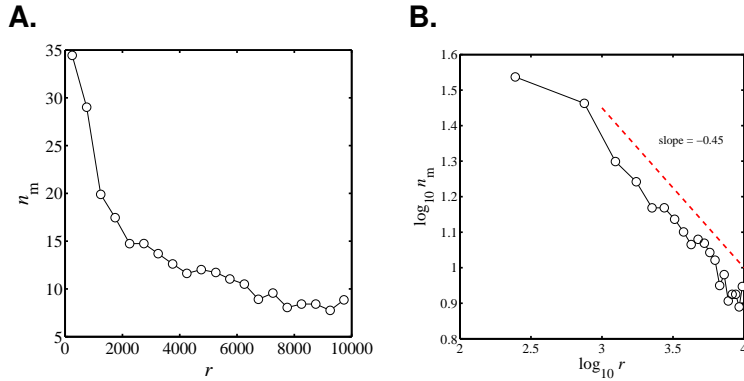
Language

Meanings
Irregular verbs
References

Frame 4/15



Word meanings



- ▶ Meaning number as a function of word rank.
- ▶ The three exponents combine within error:
 $1.2 \times 1/3 = 0.4 \approx 0.45$.

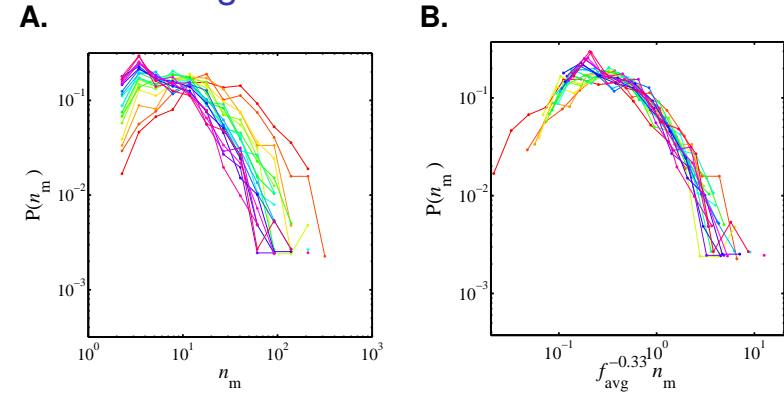
Language

Meanings
Irregular verbs
References

Frame 5/15



Word meanings



- ▶ Scaling collapse for meaning number distribution
- ▶ Each curve corresponds to approximately 500 words group according to rank (1–500, 501–1000, ...).
- ▶ With normalization

$$P(n_m) = f^{-1/3} G\left(f^{-1/3} n_m\right).$$

Language

Meanings
Irregular verbs
References

Frame 6/15



Word meanings

Further work:

- ▶ Check these scalings again
- ▶ Explore alternate data sources
- ▶ Think about why meaning number might scale with frequency.
- ▶ May be an information theoretic story.
- ▶ If we add context, we may be able to use a modified version of Simon's approach [2]
- ▶ The city story here would be that there may be many cities and towns with the same name (e.g., Springfield) with an uneven distribution in populations.

Language

Meanings
Irregular verbs
References

Frame 7/15



Irregular verbs

Cleaning up English:

“Quantifying the evolutionary dynamics of language” [1]
Lieberman et al., Nature, Vol 449, 713-716, 2007.

- ▶ Exploration of how verbs with irregular conjugation gradually become regular over time.
- ▶ Comparison of verb behavior in Old, Middle, and Modern English.

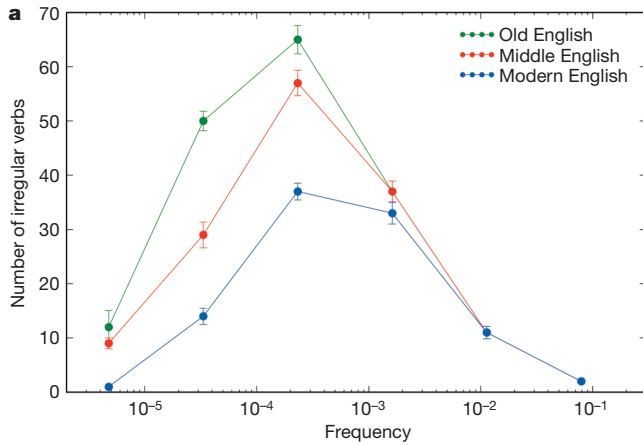
Language

Meanings
Irregular verbs
References

Frame 8/15



Irregular verbs



- ▶ Universal tendency towards regular conjugation
- ▶ Rare verbs tend to be regular in the first place

Language

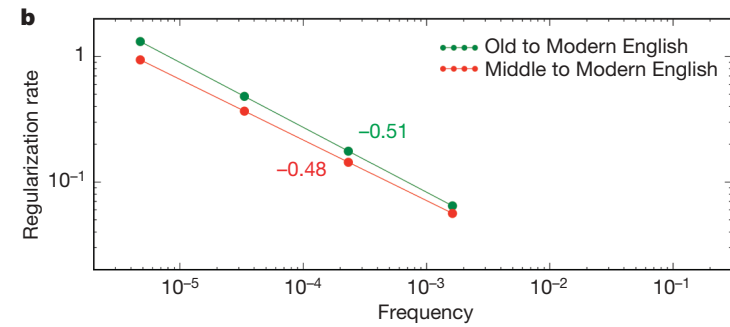
Meanings

Irregular verbs

References

Frame 9/15

Irregular verbs



- ▶ The more common a verb is, the more resilient it is to change.

Language

Meanings

Irregular verbs

References

Frame 10/15

Irregular verbs

Table 1 | The 177 irregular verbs studied

Frequency	Verbs	Regularization (%)	Half-life (yr)
10 ⁻¹ -1	be, have	0	38,800
10 ⁻² -10 ⁻¹	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie, lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, work, write	0	14,400
10 ⁻³ -10 ⁻²	arise, bake, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee, float, flow, fly, fold, freeze, grind, leap, lend, lock, melt, reckon, ride, rush, shape, shine, shoot, shrink, sigh, sing, sink, slide, slip, smoke, spin, spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, weigh, wind, yell, yield	43	2,000
10 ⁻⁴ -10 ⁻³	bark, bellow, bid, blend, braze, brew, cleave, cringe, crow, dive, drip, fans, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, writing, writhe	72	700
10 ⁻⁵ -10 ⁻⁴	bide, chide, delve, flay, hew, rue, shrive, sink, snip, spew, sup, wreak	91	300

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

- ▶ Red = regularized
- ▶ Estimates of half-life for regularization.

Language

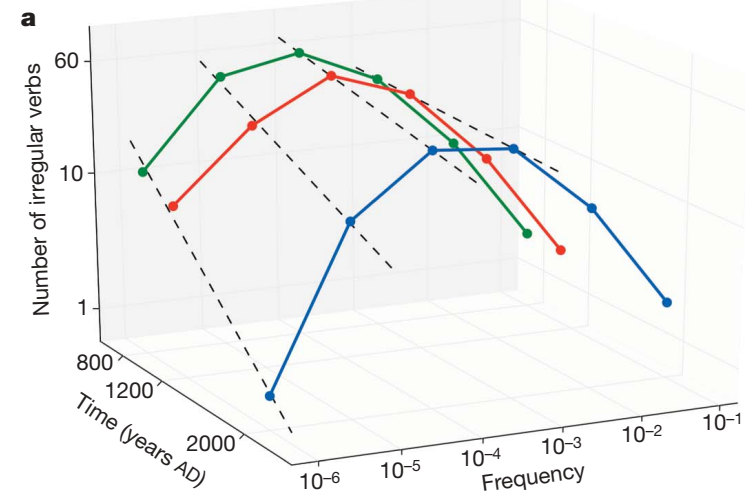
Meanings

Irregular verbs

References

Frame 11/15

Irregular verbs



- ▶ 'Wed' is next to go.
- ▶ -ed is the winning rule...

Language

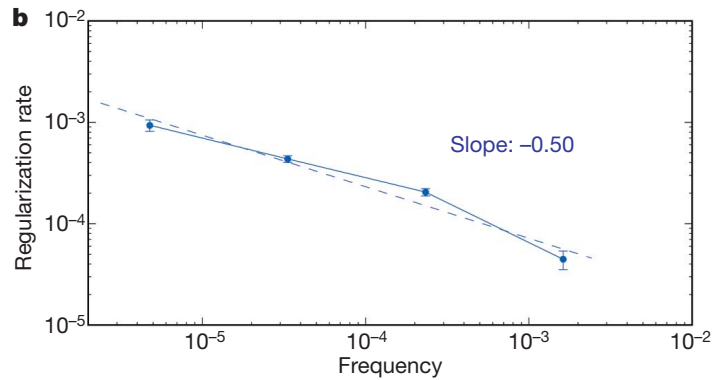
Meanings

Irregular verbs

References

Frame 12/15

Irregular verbs



- ▶ Regularization rate \propto word frequency $^{-1/2}$
- ▶ Half life \propto word frequency $^{1/2}$

Language

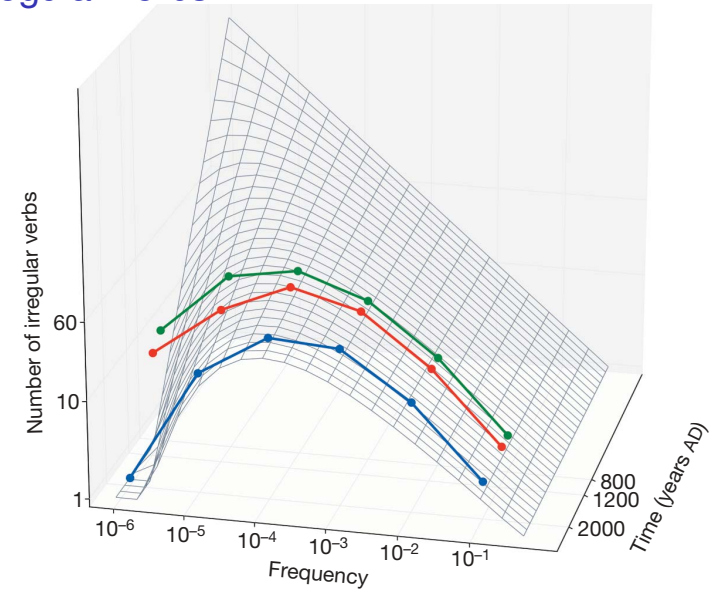
Meanings

Irregular verbs

References

Frame 13/15

Irregular verbs



- ▶ Projecting back in time...

Language

Meanings

Irregular verbs

References

Frame 14/15

References I

- E. Lieberman, J.-B. Michel, J. Jackson, T. Tang, and M. A. Nowak. Quantifying the evolutionary dynamics of language. *Nature*, 449:713–716, 2007. [pdf](#) (田)
- H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955. [pdf](#) (田)
- G. K. Zipf. *Human Behaviour and the Principle of Least-Effort*. Addison-Wesley, Cambridge, MA, 1949.

Language

Meanings

Irregular verbs

References

Frame 15/15